DOCUMENT RESUME

ED 376 138 SP 035 534

AUTHOR Hall, Kevin
TITLE Grade Expectations: The Development of a Grading Procedure and a Trial of Staff and Student Co-Assessment.
PUB DATE Jul 94
NOTE 53p.; Paper presented at the Annual Conference of the Australian Teacher Education Association (24th, Brisbane, Queensland, Australia, July 3-6, 1994).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Descriptive (141)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Elementary Secondary Education; Evaluation Criteria; Foreign Countries; Grades (Scholastic); *Grading; Higher Education; Pass Fail Grading; *Preservice Teacher Education; *Self Evaluation (Individuals); *Student Evaluation
IDENTIFIERS *University of Melbourne (Australia)

ABSTRACT

In 1989, the previously autonomous Melbourne (Australia) College of Advanced Education was incorporated into the existing Universiy of Melbourne Faculty of Education to form a new faculty, the Institute of Education. A grading procedure was developed at the Institute of Education to incorporate the College of Advanced Education's long tradition of pass/fail assessment in a new assessment policy that also incorporated the University's required grading system. Four criteria were developed for determining the appropriate number grade for each assignment; these number grades were averaged and converted to a final letter grade for the course. Staff accepted the procedure, and students rated the assessments "fair" to "very fair." A student-staff co-assessment procedure was conducted, where staff set assessment criteria and students were invited to offer self-assessments in terms of these criteria. If there was no more than one grade-level difference between the teacher's and student's assessment, the teacher's assessment was taken. If there was more than one grade-level difference, the teacher and student had a follow-up discussion to determine the grade. Of 116 assignments, students co-assessed nearly one-third. Student and teacher agreed on 35 percent. Effective student participation in co-assessment requires development of their confidence and trust. Several appendices provide additional information about assessment and grading. (JDD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 376 138

GRADE EXPECTATIONS

The development of a grading procedure and a trial of staff and student co-assessment

A paper presented
at the
Australian Teacher Education Association 24th Annual Conference,
Queensland University of Technology, Brisbane, Queensland,
3-6 July 1994

Kevin Hall

Department of Curriculum, Teaching and Learning
Institute of Education
University of Melbourne

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

K. Hall

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

SB835534

2

BEST COPY AVAILABLE

"The Educational scheme or Course established by Mr. Wopsle's great-aunt may be resolved into the following synopsis. The pupils ate apples and put straws up one another's backs until Mr. Wopsle's great-aunt collected her energies, and made an indiscriminate totter at them with a birch-rod. After receiving the charge with every mark of derision, the pupils formed in line and buzzingly passed a ragged book from hand to hand. The book had an alphabet in it, some figures and tables, and a little spelling - that is to say, it had had once. As soon as this volume began to circulate, Mr. Wopsle's great-aunt fell into a state of coma; arising either from sleep or a rheumatic paroxysm. The pupils then entered among themselves upon a competitive examination on the subject of Boots, with the view of ascertaining who could tread the hardest upon whose toes. This mental exercise lasted until Biddy made a rush at them and distributed three defaced Bibles ... This part of the Course was usually lightened by several single combats between Biddy and refractory students. When the fights were over, Biddy gave out the number of a page, and then we all read aloud what we could - or what we couldn't - in a frightful chorus; Biddy leading with a high, shrill monotonous voice, and none of us having the least notion of, or reverence for, what we were reading about. When this horrible din had lasted a certain time, it mechanically awoke Mr. Wopsle's great-aunt, who staggered at a boy fortuitously, and pulled his ears. This was understood to terminate the Course for the evening, and we emerged into the air with shrieks of intellectual victory."

From Great Expectations, (Ch. 10), by Charles Dickens, 1861

GRADE EXPECTATIONS

The development of a grading procedure and a trial of staff and student co-assessment

Kevin Hall

Department of Curriculum, Teaching and Learning
Institute of Education
University of Melbourne

INTRODUCTION

Grading is not a new or rare process in assessment. However, many challenges have to be met when grading is introduced into an environment where a Pass/Fail approach has been strongly favoured and practised over many years.

After summarising the institutional background against which this paper is set, the first major section describes the development of a grading procedure in an environment which has had a long tradition of pass/fail assessment.

In the second major section, the incorporation of staff-student co-assessment into the grading procedure is described, and the results of initial research summarised.

BACKGROUND

In 1989, the previously autonomous Melbourne College of Advanced Education was incorporated into the University of Melbourne. It was merged with the existing university Faculty of Education to form a new faculty, the Institute of Education. Although undergoing many changes of name over the years, Melbourne C.A.E. and its predecessors had had a long tradition of non-graded assessment (or, to be strictly accurate, two grades - pass/satisfactorily completed and fail/not satisfactorily completed) in Education subjects, and the practice continued for four years after

amalgamation until displaced by University policies requiring a six-point grading scale.

Principles and Practices

The commitment to non-grading was based on a number of philosophical and practical considerations which a number of former College staff, myself included, continue to hold. Examples of these considerations, some of which overlap with each other, are given below.

Those with mainly a philosophical basis can be summarised as:

Independent learning: A major aim was (and still is) to develop in our students the abilities to become independent learners. One outcome of this aim is that a degree of negotiation needs to be built-in to students' courses and, consequently, the studies undertaken might differ quite markedly from one student to another. Thus, comparing students by ranking on a common scale may be misleading because of variations in what has been studied and how it has been studied.

Competition versus co-operation: A second major emphasis has been co-operation and collaboration in teaching and learning. This involves establishing and reaching shared goals through interdependent processes in groups of various sizes. The ranking element of grading introduces a competitive atmosphere which can work against co-operation and collaboration and, further, raises practical problems about how to assess group work.

Intrinsic versus extrinsic motivation: We have endeavoured to use intrinsic rather than extrinsic motivation in our teaching, and encourage our students to do the same in their teaching. Graded assessment, through its misleading appearance of accuracy and succinctness, can become an alluring extrinsic motivator and take precedence over intrinsic aspects.

Unnecessary imposition: One reason for use of a grading system is the selection of top-ranking students for scholarships or similar awards. To impose a graded system on all students, even those who are not seeking recognition of ability to score high grades, is unnecessary. Worthy students can be identified and reported upon through means other than grades.

Major practical considerations which work against grading are:

Difficulty of measuring intangibles: Another emphasis has been, and still is, on development of reflective practice. Much reflective analysis takes place in settings and forms which are intangible and not easy to record or measure. Consequently, there is a temptation to base graded assessment on written or other tangible evidence, thus implying a devaluing of the less-tangible but often more-important forms of expression of learning such as discussion, role-play, group participation, or seminar presentation.

Similarly, the promotion of appropriate ethics, attitudes, and values is compromised because of difficulties in measuring them, and their consequent devaluing.

Potential unreliability: The more points there are on an assessment scale, the finer the discriminations need to be and thus the greater the chance of inconsistencies in assessment. It is relatively easy to distinguish between performances which meet or exceed criteria, and thus "pass", from those that "fail" to meet the criteria. It becomes more difficult to distinguish between several grades of "pass", and so the risk of error in judgement is increased.

Potential lack of validity: As subject results are not necessarily useful predictors of success in the teaching profession, we have preferred to write extensive descriptive reports as professional references for exit students. There is a danger that grades will be perceived to have more predictive accuracy than the previous "Satisfactory/ Unsatisfactory" system when this is not necessarily the case.

Inappropriateness/impracticality: In some areas or components of a course, grading is not appropriate and/or not practical. One such example is in the early stages of a teacher-education practicum program. Formative assessment in a descriptive style is more appropriate than the summative connotations of a graded result for student-teachers starting to come to grips with the realities of schools and classrooms .

Further, given the wide variety of practicum settings and experiences in terms of student-teachers' abilities and needs, pupil behaviour, availability of resources, and supervisor effectiveness, it is not practicable to expect a grading system for the practicum to meet acceptable levels of reliability without an elaborate supporting framework of communication and verification.

In developing an assessment policy to incorporate the grading requirement, we attempted to preserve as many as possible of our long-standing principles while avoiding or minimising the obvious disadvantages.

Introduction of New Policy

During 1992, in the fourth year of amalgamation, it became clear that the University was seeking to impose its policy of graded assessment on the Institute of Education. Despite a spirited defence by staff committed to the non-grading tradition, the University's will prevailed and grading commenced in 1993 in a restructured subject and its smaller companion, both running in their revised form for the first time that year. This introduction was followed by extension of the policy to all other subjects in the Institute from the beginning of 1994.

The subjects through which grading was introduced in 1993 are second-year subjects in the undergraduate Bachelor of Education (Secondary) course, a four-year "concurrent" initial teacher-education course. (This course is at present being phased out in favour of a two-year post-graduate Bachelor of Teaching degree.) The subjects are entitled "Education B - Young People, Teachers and Schools" and "Education B1 - Young People and Teachers". They are the first subjects in an Education B-C-D sequence in the second, third and fourth years of the course, Education B1 being a smaller subject tailored especially for Science students whose second year course structure does not permit the larger Education B. Education B1 students take "Education B2 - Schools" in their less-crowded fourth year. Education B and B1 students study together in Semester 1 of their second year.

In 1993, Education B/B1 had a total of 545 students enrolled and 15 staff involved in teaching. The subjects also incorporated a non-graded School Experience component. Therefore, the logistics of installing a valid and reliable graded assessment system in place of the long-standing Pass/Fail assessment in the similar predecessor subjects were quite a significant challenge in terms of changes in practice and in the number of staff and students involved.

The next section describes how the grading policy was implemented in Education B/B1 in 1993 and extended to other subjects within the Bachelor of Education (Secondary) course in 1994.

DEVELOPMENT OF A GRADING PROCEDURE

From the outset, our aims were to preserve in the new system as much of our previous philosophical position as we could, and to minimise practical difficulties.

We had to work within the confines of the University's six-point scale of:

Honours, First class	H1	80-100%
Honours, Second class - Division A	H2A	75-79%
Honours, Second class - Division B	H2B	70-74%
Honours, Third class	H3	65-69%
Pass	P	50-64%
Fail	N	0-49%

but had no other constraints explicitly imposed.

During 1992, some work had already been done on the grading of post-graduate students as staff teaching in the area found that the University's advice on grading was less than satisfactory. The post-graduate work had drawn on a University document, "Guidelines for the Use of Examiners of Theses", to form a series of statements describing the characteristics of students' work at the various grade levels. These "Guidelines", presented as passages of continuous prose, contained subjective terms such as "eminently readable", "creative sparkle", and "intellectual liveliness" scattered amongst some more-helpful criteria.

In November 1992, the Institute's Department of Curriculum, Teaching and Learning established a "Working Party on Grading" (WPOG) to formally develop the post-graduate grading policies and procedures, and in February 1993 the brief of this group broadened to include undergraduate subjects. The WPOG was able to draw on the preliminary post-graduate work in forming a basis for discussions about undergraduate implementation.

Another useful source for the WPOG was a timely article by John Biggs, "A Qualitative Approach to Grading Students", which appeared in the November 1992 issue of HERDSA News. This article describes a grading system which is based upon a series of hierarchical categories, each higher step reflecting a successively higher cognitive level. Criteria define each level and enable the grades to be used as profiles. The system is two-dimensional, recognising quality of performance as well as kind of performance, by having five cognitive stages derived from the SOLO Taxonomy (Biggs and Collis, 1982, 1986) with three levels of quality at each .

The Biggs model helped to confirm or develop two principles which were emerging from WPOG discussions between December 1992 and March 1993 - an emphasis on quality of student performance, and the use of task-focussed criteria to describe levels of performance. The proposals which emerged from the WPOG were then presented to Education B/B1 staff for further discussion. Some fine-tuning resulted and the policy and procedure which was adopted for trial during 1993 is attached as

Appendix A - "Information for Students About Assessment and Grading Procedures" and Appendix B - "Assessment and Grading - Guidelines for Staff".

This 1993 procedure, continuing with little change in 1994, required that each of the six assessable tasks in Education B (four in Education B1) be graded, and that these grades then be averaged to determine a final grade for the subject. (The School Experience component remained ungraded but, as with all practicum components in our courses, is a "hurdle" requirement - failure in the practicum means failure in the subject.)

The "Basic Criteria" in "Information for Students ..." (Appendix A, p. 1) describe increasingly higher levels of cognitive performance, expressed particularly through the ability to understand and transform source material. However, the six levels (corresponding to the University grades) are on a single continuum rather than the two-dimensional scale proposed by Biggs.

For each of the six assessable tasks, a set of "Specific Criteria" was derived from the Basic Criteria (see Appendix C - "Specific Assessment Criteria for the 'Teachers Work' Assignment" - as an example). These Specific Criteria were initially intended for use only by staff as a basis for assessing each piece of work, but it immediately became clear that they would also be of value to students. Accordingly, some staff provided photo-copies of the Specific Criteria sheets to their students, while others discussed the criteria in class.

To systematise the process of determining the appropriate grade for each piece of work, a Face Sheet was designed to record the level of achievement perceived for each of the four main criteria (see Appendix D - "Assessment Face Sheet: Schools and Their Functioning" - for an example) .

The final step in the process used to arrive at a grade for a piece of work is largely a visual one. For example, a piece of work which receives a series of four ticks down the "Excellent" column on the Face Sheet is typically graded "H1". Likewise, four ticks in line down the "Satisfactory" column typically lead to "P", two "Very good" and two "Satisfactory" to "H3", and so on. Three ticks in one column and one in another, or a wider scattering of ticks, is not as clearcut and increases the potential for subjectivity in assessment.

Although the assessment emphasis is upon quality, expressed particularly through understanding and transformation of source material, there is an assumption in the process that the four main criteria are of equal weighting. This was the subject of

some debate while the policy was being developed, but it was eventually agreed that an original, elegant, widely applicable, well-integrated and inter-related piece of work should not rate highly if it lacked any or all of the other criteria of relevance to the question or task, effort in preparation and presentation, and use of appropriate sources. Therefore, the four criteria are seen as supporting each other through inter-connection and so are accorded equal weighting.

As a check on reliability of assessments, staff exchange samples of their students' work with an "assessment partner" - a different partner for each Assignment. In most cases no adjustments have been necessary, but there have been some instances where this moderation resulted in changes to all or most of the grades in a staff member's group. The Education B/B1 Co-ordinator, Eileen Dethridge, having an overview of these changes, observed that some staff were at first reluctant to give many high grades, probably influenced by notions of a normative distribution rather than allowing the number and level of criteria met to lead to a grade. This tendency became less evident as the year progressed.

The next stage of the procedure, after grading of the six (or four) component pieces of work for each student, is to combine the component grades into a final grade.

We first contemplated using the approach adopted in the Institute's post-graduate area of looking at the profile of each students' component results in an "Examiners' Meeting" and agreeing on an appropriate final grade that reflected that profile. With 545 students and 15 staff in Education B/B1, this approach would not have been workable in terms of the large amount of time required.

Therefore, we had to resort to the use of numbers for a temporary conversion of the letter-grades to enable them to be added and averaged. The resulting average grade-marks are then converted back into a final letter-grade, the mark-range related to each particular grade being tabulated for easy reference (see Tables 2 and 3 in Appendix B - "Guidelines for Staff ..."). These mark-ranges were determined through a comprehensive series of calculations explained in Appendix E - "Determination of Ranges for Converting Grade-Number Averages to Letter-Grades".

We were concerned that the use of numbers would seduce some or many of the students, and that they would become focussed upon the quantitative rather than the qualitative aspects of the assessment process. To counteract this, we emphasise the criteria as the central focus of assessment, using the letter-grades only as a shorthand way of describing criteria met or not met, and down-play the numerical

calculations (See Appendix A - "Information for Students ...", p. 3, and Appendix B - "Guidelines for Staff...", point 1, p.1).

Frustratingly, however, University procedures require final results to be entered as a percentage mark (see table on p. 4, above) which is then converted by computer to appear as a letter-grade on the student's transcript of results.

A potential problem inherent in combining marks of different weightings is minimised by appropriately weighting each component grade before they are added and averaged. The weightings appear to have a quantitative base as they relate to the size of the tasks as described in terms of numbers of words (1000 or 2000), but there is also an important qualitative factor in that the larger tasks are also more complex and give more scope for the exercise of higher cognitive skills.

An individual record sheet format was provided for staff to adopt or adapt if they felt they needed a structure to guide them through the weighting, adding and averaging calculation steps (see Appendix F - "Memo: End-of-Year Results Procedure - Reminder").

Modifications

The system worked well over its first full year of operation. Staff accepted the procedure, and student subject evaluations at the end of the year revealed that the great majority rated assessment as "Fair" to "Very Fair". Typical supporting comments from students were "Clear criteria", " Assessment face sheet helped in diagnosis", and "Felt that I was evaluated on my ability".

The Education B/B1 policy and procedures were extended to the C-level and D-level counterpart subjects in 1994. No adjustments were made for 1994 for Education B/B1 and, while the Education Studies D staff team adopted the B/B1 policy and procedures, it appeared for a while that the Education Studies C staff team might adopt a different but related approach proposed by one of the team, John Baird, an approach which reflected the two-dimensional system advocated by Biggs (1992).

Although the alternative approach was attractive to some of the ten Education Studies C staff, it was eventually agreed that the proposal needed further development and that, for consistency in 1994, it would be better for Education Studies C to use the same approach as the B and D levels.

However, as a result of discussions about the alternative approach, some changes were made to the wording and setting-out of the Education Studies C Basic Criteria. "Relevance to question or task set'" became "Completeness and relevance", and "Evidence of effort in preparation and presentation" became "Presentation and expression". The revised setting-out included more descriptive information about the characteristics of work related to each grade-level (see Appendix G - "Extract from 'Education Studies C - General Information for Students'").

In a separate revision to minimise the potential unreliability of the ticks-in -the columns "visual" approach in converting ratings on criteria to the grade for a piece of work, the Education Studies C Assessment Face Sheet was changed to include a numbered rating scale for each criterion, the total of the assessed ratings then being matched against a series of score-ranges to determine the corresponding grade for that piece of work (see Appendix H - "Assessment Face Sheet for Assignment 1 - Classroom Data Analysis and Evaluation" as an example).

An additional modification for Education Studies C was to allow for a letter-grade for a component assessment task to be amended up or down by not more than one grade level where a staff member feels that a student's performance is not reflected appropriately by the overall number-mark for that task (see Appendix H - "Assessment Face Sheet for Assignment 1 - Classroom Data Analysis and Evaluation"). The reasons for such an amendment would be explained in the "General Comments" box.

This proviso arose out of a fear that an inflexible dependence on the numbers might sometimes produce injustices, a concern that did not arise with the Education B/B1 process because it does not use numbered rating scales for the criteria and a degree of latitude is already involved in interpreting the pattern of "ticks" and arriving at a letter-grade.

Thus, the Education B/B1 experience in 1993 served as a trial for wider implementation in 1994. The trial exposed no major problems and only a few minor ones, and so the policy and procedures were continued and extended with only fine-tuning adjustments.

Some loose ends

Despite a reasonable level of staff and student satisfaction with the system as it

stands, there are a few aspects which are of concern and which may need further consideration.

One such concern relates to an arbitrariness about some of the percentage marks that have been chosen to represent the final grades, and which are entered as the students' final results. We felt from the beginning that to use all the points on the 0 to 100 percentage scale would give very misleading impressions of accuracy and fine discrimination. Consequently, we decided to use a limited range of selected points on the 0 to 100 scale as representative of the various grade-levels.

With the grades H2A, H2B and H3 and their mark-ranges of 75-79%, 70-74%, and 65-69%, respectively, we took the mid-point of each so that a final grade of H2A is entered as 77%, H2B as 72%, and H3 as 67%. This was a relatively easy and logical decision. With the grades of H1 (80-100%), P (50-64%) and N (0-49%), however, some arbitrary decisions had to be made. They all span much larger ranges than the 5% spanned by each of the other three grades, and it was felt that they needed to be subdivided.

With the H1 grade, it was initially argued that a straight set of H1's on the component tasks should be entered as 100% but the prevailing view was that this would imply "perfect" work and that this rarely (if ever) occurred. Therefore, 97% was chosen as the highest score (requiring an average number-mark of above 5.95), with 86% representing the lower reaches of H1 (resulting from an average number-mark of between 5.55 and 5.95). (See Appendix B - "Assessment and Grading - Guidelines for Staff", Tables 2 and 3)

Likewise, the P grade-level was sub-divided into two parts with 55% representing a bare pass (an average number-mark of between 2.00 and 2.15) and 62% representing a stronger pass (an average number-mark of between 2.20 and 2.50).

Within the grade of N we decided to have three levels, drawn from Biggs (1992). Students who meet most requirements satisfactorily and who could make up the unsatisfactory component relatively easily if given another opportunity have a mark of 45% entered as their final result (an average number-mark of between 1.00 and 1.95), while students who have substantive failure and who would have to repeat all or most of the subject if given another opportunity receive a final mark of 25% (an average number-mark of between 0.05 and 0.95). Students who submit no work, or who are guilty of (in Biggs' terms) a "moral lapse" such as "gross plagiarism" receive a final result of 0%.

Although 97% and 86% both lead to an H1 grade on the student's transcript, 55% and 62% to a P grade, and 0%, 25%, and 45% to an N grade, the percentage mark is in the records as an additional indicator of level of performance, should this slightly more-specific further information be required at some later time.

A second concern is the relationship between the equal-interval 0 to 6 scale, used within the subject to convert grades on the component assessment tasks into a final grade, and the unequal-interval (ordinal?) scale used across the University for those final grades. Some staff have argued that, because the University scale decrees that a final H1 result represents a mark of between 80 to 100% (i.e., 20% of the scale), the 0 to 6 scale should reflect this so that an average mark of between, say, 4.8 and 6.0 (instead of the current 5.55 to 6.00) should earn a final H1 grade. The other mark-ranges for the remaining grades would then need to be similarly adjusted.

As a way of meeting this concern, we did consider adjusting the 0 to 6 scale so that H1 on a component assessment task would still be equivalent to a mark of 6, H2A would become equivalent to 4.5 (instead of 5), H2B would remain equivalent to 4.0, H3 would become equivalent to 3.5 (instead of 3), while P to 2 and N to 1 or 0 would retain their present relationships. This would have the effect of "squeezing" the H2A/H2B/H3 grades together and reflecting more closely the grade-intervals of the University percentage scale.

However, it was decided to retain the original 0 to 6 scale for two reasons, the first being that to use fractions (6.0, 4.5, 4.0, 3.5, 2.0, 1.0, 0) would make calculations a little more complex (some staff are slightly fazed by the current process mainly using whole numbers, even with tables to assist in calculation!).

The second reason is based on the view that we are starting from a valid base of a criterion-referenced assessment process, and we are projecting the data which are produced by this process - the final grades - on to the University scales. This "upward" process is, we feel, a preferable alternative to starting with the University scales (which appear to have no valid base) and making major "downwards" adaptations to fit our process to them.

These concerns may be further considered when reviewiing policy and practices for 1995.

The next section discusses a student-staff co-assessment procedure conducted within the Education B/B1 framework in 1993 and 1994, and extended as well to Education Studies C in 1994.

CO-ASSESSMENT

A definition

"Co-assessment" is used here to refer to a situation where student and teacher participate in assessment as a joint effort. Elsewhere (Hall 1981, 1992), I have distinguished co-assessment from student self-assessment and teacher-assessment as follows:

Student self-assessment is the case where a student assesses herself or himself, on the basis of criteria which she or he has selected, the assessment being either for the student's private information or for communication to the teacher or others. The two critical factors for "self"-assessment are that the student not only carries out the assessment but also selects the criteria on which the assessment is based. Whether the assessment outcome is to be kept private or made public is of less importance.

Similarly, teacher-assessment is where the teacher both selects the criteria and carries out the assessment of the student.

Any situation where the teacher and student share in the selection of criteria and/or the carrying-out of the assessment is more accurately termed "co-assessment". By these definitions, many instances of what are referred to in the literature as "student self-assessment" involve teacher-set criteria and therefore are more accurately termed "co-assessment".

In the co-assessment situation being described here, the criteria had been set by staff, and students were invited to offer their own assessment in terms of these staff-set criteria.

Purposes

Several purposes underlie the introduction and use of this co-assessment process. One is to assist the student-teachers in making the role-change from being a student to being a teacher, a second is to provide insights into the assessment process

which may be of use to them in assessing their own students, and a third is to provide a skill-development step towards self-assessment.

Making the role-change from being a student and responsible for one's own learning to being a teacher and responsible for the learning of others is difficult for some students. If teacher-education staff dominate the staff-student relationship by over-playing the roles of "expert" and decision-maker, then students have less space and less incentive to develop as independent learners. To open-up the assessment process to co-assessment is one way of encouraging and fostering this independence and accompanying responsibility.

Assessment is a complex process and a crucial element in education, but many student-teachers go through their teacher-education courses without much study of or practice in this important area. To be involved in their own assessment is one way of helping students to learn about what assessment is and how to do it.

Self-assessment, and independent learning in general, requires particular skills. As defined above, the two critical factors for self-assessment are that the student not only carries out the assessment but also selects the criteria on which the assessment is based. Co-assessment, by involving the student in the process, offers a stepping-stone towards self-assessment where the student can develop her or his own criteria and carry out her or his own assessment.

The purpose of the analysis which follows is to illuminate the workings of the co-assessment process in order to facilitate improvement. The analysis focusses upon the level of participation and the degree of staff-student agreement, but these are simply pointers to other aspects of the process.

For example, if the proportion of students taking the opportunity to self-assess is low, we would need to look at the way in which the process is presented and the advantages and disadvantages that students perceive as a result of participation. If the level of staff-student agreement is low, we would probably need to look at the criteria in terms of their relevance and explicitness, and at the ratings scales that apply to them.

The process

The invitation to co-assess was offered initially to 33 Education B students during Semester 2, 1993, on each of the final two assignments for the year - the fifth and

sixth pieces of submitted work in the subject. The first of these required a response of 2000 words or equivalent, and the second a response of 1000 words or equivalent.

The invitation was offered again to 25 Education B/B1 students in Semester 1, 1994, on each of the first two assignments for the year. These assignments were of 2000 words and 1000 words, or equivalent, respectively.

A special double-sided version of the Assessment Face Sheet was used to allow for the co-assessment option (see Appendix I - "Education B 1993 [Kevin's Groups], Face Sheet for Schools and Their Functioning"). On the back was a box headed "Student's Assessment", reflecting the standard "Staff Assessment" box on the front of the Face Sheet. I promised the students that I would not look at the "Student Assessment" box until after I had arrived at my assessment and recorded it in the "Staff Assessment" box.

If they had recorded their own assessment, folding-over of the Face Sheet put the two assessments side-by-side for easy comparison. If the two assessments agreed, then the system was working well. If they did not agree, my initial position was that there would need to be follow-up discussion in each case about why our views differed and to negotiate an agreed grade, while reserving my right to make the final decision as I believe must be the case in a credentialling course (Hall 1992).

However, being initially unsure of the number and extent of such differences that might arise, I eventually decided to take the more cautious approach of taking my assessment as the one to be recorded for the assignment if there was no more than one grade-level difference between my assessment and the student's assessment, and only following-up with discussion where there was more than one grade-level difference. (As the accompanying data shows, I tended to rate them more highly than they did themselves, so complaints about my making the final decision were unlikely!)

There were 13 cases over the four assignments where there was more than one grade-level difference (see Appendix J - "Analyses of Staff and Student Assessments", Table 2), my assessment being higher than the student's in 12 of these cases. There were only three cases where the student's assessment was higher than mine, and in two of these there was only one grade difference. In all of these cases, the students accepted my grade without any evident objection.

The outcomes

It needs to be kept in mind that this is a report of work in progress. The data so far is limited, and derived from two groups of students at different stages of the subject so that in some aspects it cannot be aggregated. The study will continue in 1994 to provide at least one full year's data.

In summary, the following points can be made regarding the four assignments:

In terms of **participation**,

- Of a total of 116 assignments, nearly one-third (31.9%) on average were co-assessed (45.5% in late 1993 and 14% in early 1994).

In terms of the **overall grades** for the Assignments,

- The students and I agreed in 35.1% of the co-assessments.

- I assessed them at a higher grade than they did in 56.8% of the cases and lower in 8.1% .

- Staff/Student agreement was most frequent at the "H2B" level, "H2A" and "H1" levels, in that order.

- Where there was difference, the most frequent staff(student) combinations were "H2A(H3)" and "H2A(H2B)".

In terms of **Specific Criteria**,

- The students and I agreed in 35.1% of the instances (coincidentally exactly the same level of agreement as that on overall grade).

- I rated them higher than they did in 54.1% of the instances and lower than they did in 10.8%.

- Staff/Student agreement was most frequent on "Effort", followed by "Understanding" and "Sources". "Relevance" was the criterion of least agreement. (Note: "Effort" is somewhat of a misnomer. It is mainly concerned with preparation and presentation.)

- Where there was difference in the rating of Specific Criteria, the most frequent staff(student) combinations were "Excellent (Very good)"and "Excellent (Satisfactory)".

A more-detailed summary follows, with some questions and comments added in italics. Appendix J - "Analysis of Staff and Student Co-Assessment, Education B/B1 1993-94" contains the data from which these summary points are drawn, and relevant Table numbers are given.

(Because the 1993 data present a different picture to the 1994 data in some aspects, they are dealt with separately in many of the following points.)

Participation rates (see Table 1) were as follows:

- 58 students were in the groups invited to co-assess (44 females, 14 males). They submitted a total of 116 assignments, of which 37 (31.9%) were co-assessed. However, this figure masks a large difference between the 1993 and 1994 co-assessment proportions - 45.5% and 14%, respectively.
  (A likely explanation for this difference is that a greater proportion of the 1993 students, having completed one semester and four previous pieces of work in the subject, felt more comfortable about participating in co-assessment than did the 1994 students in their first semester and tackling their first two pieces of formal submission.)

- 15 students co-assessed on each of the first two of the four assignments concerned (Semester 2, 1993), and 2 students and 5 students respectively on the third and fourth assignments (Semester 1, 1994).
  (The increase from 2 to 5 co-assessments between the first and second 1994 assignments supports the "increasing comfort" suggestion in the point above.)

- 8 of the 1993 students offered their assessment on both assignments, leaving 7 who offered only on the first and 7 who offered only on the second. Of the 1994 students, 2 students co-assessed on both assignments and 3 others on only the second assignment.

- On the 1993 assignments, females were over-represented in co-assessment - 28 of the 30 pieces of work (93.3%) were submitted by females (81.8% of the class). However, they were under-represented in 1994 - 57.1% of the co-assessed pieces of work, although 68% of the class.

The degree of staff-student agreement on **overall grade** (see Table 2) was as follows:

Perfect agreement occurred in 13 instances out of the 37 (35.1%) - 3 at the "H1" level, 4 at "H2A" and 6 at "H2B".
(Why is agreement less likely at the top levels?)

- I assessed them at a higher grade than they did in 21 instances (56.8%) and lower in 3 instances (8.1%).

- There was only one grade level difference between staff and student assessments in 11 instances (29.7%), two grade levels difference in 12 instances (32.4%), and 3 grade levels difference in 1 instance (2.7%).

- Where there was difference, the most frequent staff(student) combinations were "H2A (H3)" - 6 instances, and "H2A(H2B)" - 5 instances.

- No consistent gender differences emerged apart from a strong tendency for the few males involved to under-assess themselves in comparison to my grade. We agreed in one case of the five and they under-assessed in the other four. In the rank order of grades for each assignment, the males were in the middle to lowest positions.
(That I tended to give them higher grades than they did themselves may suggest that I was favouring males, but this is not the case as I do not look at the student's name until I have finished reading the assignment and forming an assessment.)

The degree of staff-student agreement on **Specific Criteria** (see Table 3) was as follows:

- Perfect agreement occurred in 52 instances (35.1%) out of a possible 148 (i.e., 4 criteria on each of 37 co-assessed submissions). Of the remaining 96 cases, I rated the students more highly than they did themselves in 80 instances (54.1%), and lower than they did in 16 instances (10.8%).
(In the cases of difference, my "Excellent" and "Very good" assessments tended to be higher than the students", and my "Satisfactory" and "Unsatisfactory" lower than the students'.)

- In 75 instances (50.7%), there was a difference of one rating level between my assessment and the student's assessment, in 20 (13.5%) there was a

difference of two rating levels, and in 1 (0.7%) a difference of three levels. Differences were most likely in cases where I had rated the students as "Excellent" (68 of the 148 cases - 45.9%).

- Taking each of the four Specific Criteria separately, perfect agreement between staff and student assessments occurred more often on the "Effort" (20 instances) and "Understanding" criteria (14 instances) than on "Sources" (11 instances) and "Relevance" (7 instances).
  (Is it that some criteria are more difficult to assess than others, or that they are less clearly-defined?)

- The most common Staff(Student) rating combinations occurred at the "Excellent (Very good)" level - 47 cases (31.8%), "Very good (Very good)" - 28 cases (18.9%), "Excellent (Satisfactory)" - 20 cases (13.5%), and "Excellent (Excellent)" - 18 cases (12.2%).

- Again, reflecting the Overall Grade data, consistent gender differences do not emerge apart from a tendency for males to under-assess their criteria ratings by comparison with mine. Of the 20 cases (5 pieces of co-assessed work x 4 criteria) 13 were under-rated, we agreed on six, and in one case the student suggested a higher rating than mine.

Regarding the **grade distributions** of co-assessed and staff-assessed students, the data is inconsistent. Comparing the grades of those who co-assessed with the grades of those assessed by staff only (see Tables 4A and 4B), the following points can be made:

- On the 1993 data, most co-assessing students received a grade of "H2A" or "H2B" (36.7% at each level), while most staff-assessed students received "H1" or "H2A" grades (30.6% and 27.6% respectively).

  On the 1994 data, most co-assessing students received "H2A" grades (62.5%) while most staff-assessed students received "H2B" grades (38.1%).

Follow-up

In 1994, besides my Education B/B1 students, I am offering the co-assessment opportunity to my Education Studies C students (two groups totalling 55 students), and to an Education Studies D specialist elective group of three students. However,

at the time of writing, the first C and D level students submissions had not been submitted and so could not be included in these analyses.

My policy will continue to be that discussion and negotiation will only occur if there is a difference of more than one grade, and that, if consensus is not then reached, my decision will prevail. (In the context of co-assessment, with its co-operative ideology, this may sound high-handed and hypocritical. However, I see this discretion as a last resort, to be used only after co-operative and consensus approaches have been fully explored.)

At the end of 1994, as part of our regular program evaluations, I will gather feedback from all students in my groups about why they did or did not participate in the co-assessment process and the perceived advantages and disadvantages. This should illuminate the accumulating statistical data.

Further, the larger amount of data should permit some deeper statistical analysis of correlations.

What does the research literature say?

The research literature is relatively sparse and widely-scattered. It is blurred by overlapping terminology and is drawn from all levels of education, primary to tertiary. Nevertheless, some common guiding principles can be identified.

In an earlier literature review of student self-assessment (Hall, 1981), the following points emerged:

- the small amount of research in student self-assessment and related areas over a surprisingly long period (back to the 1920's)
- a confusion between co-assessment and student self-assessment
- the necessity for skill development for effective self- or co-assessment
- beneficial effects particularly on students attitudes and perhaps also on achievement.

These points, and two additional ones, are used as a framework to summarise some recent research. Two articles by Boud and Falkichov reviewing research on self-assessment in higher education have been of particular use in this brief overview.

Lack of research

Student self-assessment and related areas still seem to attract relatively little attention. In a recent review of research on student self-assessment in higher education, Falchikov and Boud (1989: 395) have commented that "it is surprising that, until 1989, no major review of the literature seems to have been undertaken". However, in a related article (Boud and Falchikov, 1989:530), they note that "there has been an upsurge in interest in self-assessment in the past ten years" and identify two main reasons for this,

> "... one primarily educational, the other often expedient. Firstly ... a principled desire on the part of teachers for learners to take greater responsibility for their own learning ... Secondly ... a practical need to develop assessment procedures which are a more effective use of resources through using students more and teachers less".

Terminology

The confusion of terminology and the practices to which it is applied continues still. Boud and Falkichov (1989: 529) state that

> "Many studies which describe themselves as studies of self-assessment do not involve students in the selection of criteria and simply ask them to rate themselves according to some pre-established scale"

and

> "Where students are involved in making judgements of their work without a concomitant involvement in establishing criteria, this is commonly referred to as self-marking."

This is a form of what I prefer to call "co-assessment", as headed and defined at the beginning of this section, a term adapted from Bloch (1977). However, most literature references appear with the prefix "self- ..." and it is from such sources that I have drawn. My view is that the same general principles apply, whether self- or co-assessment, the difference by my definition being the degree of student involvement.

Skill development

The need for a developmental process is recognised by Rudd and Gunstone (1993:20). They define four overlapping stages in a teacher's role in developing self-assessment skills in students: "The teacher as instructor", taking a dominant role in

shaping what is to be done and how; "The teacher as coach", moving towards a form of partnership but with the teacher still more dominant than the student; "The teacher as counsellor", a partnership but with the student more dominant and the teacher available for advice; and "The teacher as delegator", where the teacher delegates the Stage 1 role and the student is responsible for applying previous learning.

Jensen and Loacker (1988:130) also recognise a developmental process: "As students develop their understanding of the role criteria play in their education, they are increasingly able to take more responsibility for their own learning".

Taking the 1993/94 participation rates reported in the data above, the 1993 students had previously completed four other assignments before being invited to co-assess. Therefore, they may have been more willing to become involved because they had a greater understanding of what was expected than did the 1994 students, who were facing the first two of their six assignments.

Boud and Falkichov (1989:425) claim that "there is particularly a lack of studies on the influence of practice on self-marking", but, besides practice, there may be at least two other factors contributing to skill development in self-assessment or co-assessment (at least in higher education) - expertise and ability.

Falkichov and Boud (1989:425) found:

> "Senior students taking introductory courses appear not to self-assess significantly better than do first-year students. Students in advanced courses, however, where self-assessment appears to be particularly accurate, are also students often classified a seniors. Thus we must conclude that expertise within a particular field is more influential than is seniority or duration of enrolment."

With regard to ability, after making the point that their review shows "no consistent tendency to over- or underestimate performance", and that " some students in some circumstances tend towards one direction, others in the same or different situations towards the other", Boud and Falkichov (1989:543) note that "the review also points to the ability of self-assessors as a salient variable, with the more able students making more accurate self-assessment than their less able peers."

With respect to reliability and the correlation between student and staff assessments, my finding that higher-graded students tend to give themselves a lower grade than mine and that lower graded students tend to give themselves a higher grade parallels that of Boud and Falkichov's (1989:541), namely

> "The general trend in these studies suggest that high achieving students tend to be realistic and perhaps underestimate their performance while low achieving students tend to overestimate their achievements probably to a greater extent than the underestimation".

However, an important point to keep in mind is that there is an assumption that it is the staff member's grade which is "correct". As Boud and Falkichov (1989:536) put it,

> "At the simplest level of performance where we can assume teachers to be experts and students to be novices there is little difficulty in adopting this as a valid working assumption. However, as students progress to higher levels of sophistication and begin to apply their knowledge and understanding to increasingly complex professional questions which begin to fall outside their teachers' immediate area of competence, then the assumption begins to be less valid.
>
> In addition we need to recognise that teachers have limited access to the knowledge of their students and in many ways students have greater insights into their own achievements ... Furthermore, teachers and students may have different perspectives and differing ideas about what is important."

Boud and Falkichov (1989:537) also point out that

> "In most studies greater numbers of student marks agree rather than disagree with staff marks. ... Not surprisingly, there is a much greater chance of agreement between staff and students when a five point scale is used rather than percentages."

Boud and Falkichov (1989:543) note that "Studies of gender differences remain inconclusive", and the co-assessment data reported here is similarly unclear.

Beneficial effects

Falkichov and Boud (1989:427) are of the opinion that "Self-assessment can be a valuable learning activity, even in the absence of significant agreement between student and teacher, and can provide positive feedback to the student about both learning and educational and professional standards."

The effects of self- or co-assessment are often not explained in the studies in this area (or perhaps not investigated?), but the tenor of the reports suggests that, as found in my earlier review, attitudinal outcomes are predominant.

Students' trust and confidence

Taking the previous point further, there is an attitudinal prerequisite for co-assessment - student's must have trust and confidence in the process, and must be willing to participate in co-assessment. They could be coerced or otherwise persuaded but, unless they feel that they have some real power in the process, their participation is likely to be mechanical and of little contributive value.

A very important element in engendering in students the feeling that assessment power is shared is the general approach taken by the staff member with the students concerned. If a feeling of openness and trust can be developed across the range of activities that the staff member and students are involved in, then the students should have more confidence that the co-assessment process will be carried out in the same way.

As noted above, I feel that the higher proportion of students willing the co-assess in 1993 was due to the fact that we had already worked together for more than a semester, whereas the 1994 students and I were still developing our relationship in our first semester together.

Rudd and Gunstone (1993:4) note the importance of "the need for time, the importance of embedding self-assessment in learning contexts seen as part of the normal curriculum, the need for trust between teacher and student".

As further data from my research accumulates over a full year, it will be interesting to see if the participation rates increase, and if the level of agreement increases. In addition, if there are such increases, will they extend beyond year levels (e.g., will there be greater levels of agreement in D level subject co-assessment than in C level and B level)?.

Summary

We still do not know much about self-assessment and related areas. The various models and the terminology describing them needs clarification. It is obvious that there is a developmental process towards effective self- or co-assessment but the

effects of variables such as practise in assessing one's own work, expertise in a particular area, general ability, or gender are not clear. It is clear, though, that students confidence and trust need to be obtained if effective participation is to be realised.

Acknowledgements:

Thanks are due to my colleagues Bill Stringer, Gaell Hildebrand, Roger Hewitt, Eileen Dethridge and John Baird for their comments on the draft of the first section of this paper (Development of a Grading Procedure).

BIBLIOGRAPHY

Biggs, John (1992), "A Qualitative Approach to Grading Students", HERDSA News, 14 (3), 3-6

Biggs, John, and Collis, Kevin, (1982), Evaluating the Quality of Learning: the SOLO Taxonomy, New York: Academic Press

Biggs, John, and Collis, Kevin, (1986), "Using the SOLO Taxonomy", SET - Research Information for Teachers, No. 1, Item 3(A)

Bloch, J.A., "Student evaluation in 'individualised' science programs", Studies in Educational Evaluation, Summer 1997, 3(2), 95-107

Boud, David, and Falchikov, Nancy, (1989), "Quantitative studies of student self-assessment in higher education: a critical analysis of findings", Higher Education, 18, 529-549

Falchikov, Nancy, and Boud, David, (1989), "Student Self-assessment in Higher Education: A Meta-Analysis", Review of Educational Research, Winter, 59(4), 395-430

Jensen, Patricia, and Loacker, Georgine, (1988), "The power of performance in developing problem solving and self assessment abilities", Assessment and Evaluation in Higher Education, Summer, 13(2), 128-150

Hall, Kevin, (1981), Student Self-Evaluation: A Review of the Literature and Four Case Studies, Unpublished M. Ed. Studs. Project: Monash University

Hall, Kevin, (1992), Co-Assessment: The Bridge Between Student Self-Assessment and Teacher Assessment, Paper presented at 12th Annual International Seminar for Teacher Education, University of New England, Armidale, N.S.W., April 24-30 and published in Building Bridges in Teacher Education - Proceedings of the 12th Annual International Seminar for Teacher Education, University of New England:1994

Rudd, Telsa, and Gunstone, Richard, (1993), Developing Self-Assessment Skills in Grade 3 Science and Technology: The Importance of Longitudinal Studies of Learning, Paper presented at the Annual Meetings of the National Association for Research in Science Teaching, Atlanta GA, U.S.A., April 15-18, and the American Educational Research Association, Atlanta GA, U.S.A., April 12-16, ED358103

APPENDICES

A Information for Students About Assessment and Grading Procedures

B Assessment and Grading - Guidelines for Staff

C Specific Assessment Criteria for the Teachers' Work Assignment

D Assessment Face Sheet - Schools and Their Functioning

E Determination of Ranges for Converting Grade-Number Averages to Letter-Grades

F Memo: End-of-Year Results Procedure - Reminder

G Extract from "Education Studies C - General Information for Students"

H Assessment Face Sheet - Assignment: Classroom Data Analysis and Evaluation

I Education B 1993 (Kevin's Groups) - Assessment Face Sheet: Schools and Their Functioning

J Analyses of Staff and Student Assessments - Comparisons Across Four Assignments

UNIVERSITY OF MELBOURNE INSTITUTE OF EDUCATION

Bachelor of Education (Secondary) Course

EDUCATION B/B1, 1993

APPENDIX A

INFORMATION FOR STUDENTS ABOUT ASSESSMENT AND GRADING PROCEDURES

ASSESSMENT CRITERIA

A criterion-referenced system of assessment will be used in Education B/B1. That is, criteria are set which describe the various levels of achievement or performance that can be reached and your work is matched against these criteria to determine your appropriate grade.

You will be advised of any specific assessment criteria for each task, but the basic criteria which will be used for each of the 6 grade levels used by the University of Melbourne are as follows:

GRADE	BASIC CRITERIA (Words in bold type indicate additions or refinements at each successive level.)
P	Relevance to question or task set. Evidence of effort in preparation and presentation. Appropriate sources located and used. Understanding of the material being presented. *BUT* *Little if any transformation of sources.* *Description rather than analysis and interpretation .* *Listing rather than inter-relating or integrating.*
H3	*All "P" criteria and some "H2B" criteria met.*
H2B	Relevance to question or task set. Evidence of effort in preparation and presentation. Appropriate sources located and used well. **Sound** understanding of the material being presented. Selectivity and judgement shown in what is important. **Transformation of sources by analysis and interpretation or inter-relating or integrating.** **All parts relate well to form a coherent whole.**
H2A	*All "H2B" criteria and some "H1" criteria met.*
H1	Relevance to question or task set. Evidence of effort in preparation and presentation. Appropriate sources located and used very well. **Thorough** understanding of the material being presented. Selectivity and judgement shown in what is important. Transformation of sources by analysis and interpretation or inter-relating or integrating. **High level of abstract thinking and synthesis.** **High level of originality, elegance, or generalisation or application to other contexts.** All parts relate well to form a coherent whole. **Overall an outstanding piece of work.**
N	*One or more criteria for a "P" grade not met* *OR* *No work submitted.*

DETERMINATION OF FINAL GRADE

The grades you receive on the various assessment tasks during the year in Education B/B1 will be combined to reach a final grade in the following way:

The letter-grades you received for each task will be converted temporarily to number marks so that they can be added and averaged. A grade of H1 will be taken to be equivalent to 6, H2A to 5, H2B to 4, H3 to 3, P to 2, and N to 1. A mark of 0 will be recorded if no work is submitted.

If the assessment task has a requirement of 1000 words or equivalent, your mark will remain as a mark out of a possible 6. If it is 2000 words or equivalent your mark out of 6 will be multiplied by a weighting of 2 to give you a mark out of a possible 12, and if it is 3000 words or equivalent your mark out of 6 will be multiplied by a weighting of 3 to give you a mark out of a possible 18.

Therefore, the 9000 words required for Education B will lead to a possible overall total of 54 (i.e., 6 x 9), while the 6000 words required for Education B1 will lead to a possible total of 36 (6 x 6).

The marks you received for each of the graded assessment tasks will be added together to give you a total mark out of the possible 54 (for Education B students) or 36 (for Education B1 students), as the case may be. This total mark will then be divided by 9 (Education B) or 6 (Education B1) to bring it back to an average mark on the 6 to 0 scale.

For example, take an Education B student who has scored the following on each of the six* graded assessment tasks:

Task	Number of words	Student's letter-grade	Equivalent score out of 6	Weighting	Student's recorded score for that task
1	2000	H2B	4	x2	8 (out of a possible12)
2	1000	H2A	5	x1	5 (out of a possible 6)
3	2000	H3	3	x2	6 (out of a possible 12)
4	1000	H1	6	x1	6 (out of a possible 6)
5	2000	H3	3	x2	6 (out of a possible 12)
6	1000	H2A	5	x1	5 (out of a possible 6)
		End of year total:		x9	36 (out of a possible 54)

The end-of -year total of 36 is divided by 9. This gives an average of 4.0, which is equivalent to a final grade of H2B.

However, staff are required to enter final marks in the form of percentage marks which the computer then converts to the letter grade that you receive, as follows:

A final mark of	80-100%	receives a grade of	H1
	75-79		H2A
	70-74		H2B
	65-69		H3
	50-64		P
	0-49		N

Therefore, to enable staff to enter a percentage mark as the final result, the following conversion table will be used:

A final average mark of ...	*... will be recorded as a final percentage mark of ...*	*... which will appear on your transcript as a final grade of ...*
6.00 5.55 to 5.95	97) 86)	H1
4.55 to 5.00 to 5.50	77	H2A
3.55 to 4.00 to 4.50	72	H2B
2.55 to 3.00 to 3.50	67	H3
2.20 to 2.50 2.00 to 2.15	62) 55)	P
1.00 to 1.95 0.05 to 0.95 0.00	45) 25) 0)	N

Note, though, that all of the component assessment requirements - attendance, participation, School Experience, and 6 (Education B) or 4 (Education B1) assessment tasks - must be passed to gain a final grade of P or higher. That is, failure in one or more of the component items will lead to a final grade of N, even if the final average over the other items is 2.00 or higher.

(At the end of Semester 1, a Committee of Review will consider the cases of students who have failed any of the assessment tasks to that point.)

IMPORTANT!

It is important to keep in mind that the criteria are the central focus of the assessment procedures. The letter-grades are simply a shorthand way of describing which criteria you have achieved, and the number-marks have no importance on themselves - they are only a temporary conversion mechanism to calculate a final grade.

BEST COPY AVAILABLE

UNIVERSITY OF MELBOURNE INSTITUTE OF EDUCATION

Bachelor of Education (Secondary) Course

EDUCATION B/B1, 1993

APPENDIX B

ASSESSMENT AND GRADING - GUIDELINES FOR STAFF

1. When providing information for the students about each graded assessment task, provide details of specific criteria (if any) applying to each of the 7 points on the H1 to N scale. These specific criteria should be based on the general criteria listed in the "Information for Students About Assessment and Grading Procedures" handout.

   In discussion with students about assessment, emphasise the criteria rather than the letter-grades or number-marks.

2. When students submit their work for assessment, give each piece of work an initial grade of H1, H2A, H2B, H3, P, or N, according to the criteria met or not met.

3. At the end of the year (or progressively during the year if you prefer), for each graded assessment task that a student has done, temporarily convert the grade to a number mark for the purpose of adding and averaging to arrive at a final mark and grade.

   H1 will be equivalent to 6, H2A to 5, H2B to 4, H3 to 3, P to 2, and N to 1. If no work is submitted, record 0. Use whole numbers, e.g., 3 or 4, not 3.5.

4. Weight these number marks where necessary. Taking a task requiring 1000 words or equivalent as having a base weighting of 1, the marks out of 6 for 2000 and 3000 word tasks need to be multiplied by a weighting as follows in Table 1:

   TABLE 1

If the task requires ...	*... multiply the mark out of 6 by ...*	*... to give a mark out of a possible ...*
(1000 words or equivalent	1	6)
2000 words or equivalent	2	12
3000 words or equivalent	3	18

5. For each student, record the weighted mark for each of the 6 graded assessment tasks (4 for Education B1) completed during the year.

6. At the end of the year, add the 6 weighted marks (4 for B1) to give a total for the year out of a possible maximum 54 (or 36 for B1).

7. Using Table 2 (over the page), find the student's total mark out of 54 (or 36) in the "Final Total Mark" column and read across to find:

   the corresponding average,

   the percentage mark to be entered as the student's final result (if all assessment tasks have been satisfactorily completed - see Step 8), and

   the grade that will eventually appear on the student's transcript of results.

TABLE 2*

EDUCATION B			
Final total mark	Average (Total divided by 9)	% Mark	Grade
54	6.00	97	
53	5.89	86	H1
52	5.78		
51	5.67		
50	5.56		
49	5.44	77	H2A
48	5.33		
47	5.22		
46	5.11		
45	5.00		
44	4.89		
43	4.78		
42	4.67		
41	4.56		
40	4.44	72	H2B
39	4.33		
38	4.22		
37	4.11		
36	4.00		
35	3.89		
34	3.78		
33	3.67		
32	3.56		
31	3.44	67	H3
30	3.33		
29	3.22		
28	3.11		
27	3.00		
26	2.89		
25	2.78		
24	2.67		
23	2.56		
22	2.44	62	P
21	2.33		
20	2.22		
19	2.11	55	
18	2.00		
17	1.89	45	N
16	1.78		
15	1.67		
14	1.56		
13	1.44		
12	1.33		
11	1.22		
10	1.11		
9	1.00		
8	0.89	25	
7	0.78		
6	0.67		
5	0.56		
4	0.44		
3	0.33		
2	0.22		
1	0.11		
0	0.00	0	

EDUCATION B1			
Final total mark	Average (Total divided by 6)	% Mark	Grade
36	6.00	97	H1
35	5.83	86	
34	5.67		
33	5.50	77	H2A
32	5.33		
31	5.17		
30	5.00		
29	4.83		
28	4.67		
27	4.50	72	H2B
26	4.33		
25	4.17		
24	4.00		
23	3.83		
22	3.67		
21	3.50	67	H3
20	3.33		
19	3.17		
18	3.00		
17	2.83		
16	2.67		
15	2.50	62	P
14	2.33		
13	2.17	55	
12	2.00		
11	1.83	45	N
10	1.67		
9	1.50		
8	1.33		
7	1.17		
6	1.00		
5	0.83	25	
4	0.67		
3	0.50		
2	0.33		
1	0.17		
0	0.00	0	

8. If there are no "fail" scores (1 or 0) amongst the graded assessment tasks and all of the ungraded tasks (attendance, participation, School Experience) have been satisfactorily completed, the percentage mark derived in Step 7 from Table 2 can be entered as the student's final result.

If one or more of the assessment tasks, graded or ungraded, has not been satisfactorily completed then enter a final result of 45%, 25%, or 0%, even if the student's final average score is 2.00 or higher.

A final result of 45% should be used for a case where most requirements have been satisfactorily met and the unsatisfactory component could be made up relatively easily if the student was given another opportunity.

A final result of 25% should be used for a case of substantive failure where the student would have to complete all or most of the subject, if given another opportunity.

A final result of 0% should be used for a case where no work was submitted or a student has suffered a "moral lapse" and is found guilty of gross plagiarism.

Note: Table 2 is derived from the relationships shown in Table 3 below.

TABLE 3

A final average mark of ...	*... will be recorded as a final percentage mark of ...*	*... which will appear on the student's transcript as a final grade of ...*
6.00 5.55 to 5.95	97) 86)	H1
4.55 to 5.00 to 5.50	77	H2A
3.55 to 4.00 to 4.50	72	H2B
2.55 to 3.00 to 3.50	67	H3
2.20 to 2.50 2.00 to 2.15	62) 55)	P
1.00 to 1.95 0.05 to 0.95 0.00	45) 25) 0)	N

BEST COPY AVAILABLE

EDUCATION B/B1, 1993

SPECIFIC ASSESSMENT CRITERIA FOR THE "TEACHERS' WORK" ASSIGNMENT

BASIC CRITERIA	SPECIFIC CRITERIA
Relevance to question or task set.	**In option 1**, there are **two parts** to the discussion of the statement - a "**now**" and a "**should be**". **In option 2**, the students should "**list and discuss**" the approaches/perspectives/ philosophies that they prefer, **and explain their preference(s)**.
Evidence of effort in preparation and presentation.	**Spelling, expression, legibility, neatness, use of conventions** for referring to sources, etc., fall under this criterion.
Appropriate sources located and used.	**The specified sources are "the Teachers' Work section of the resource book, material covered in lectures and seminars" and "other resources"**. (Note: If the student has had primary and/or secondary School Experience, this should be a potential source. However, they should not be disadvantaged if they have not been placed.) The more of these that are used, and the more specific the references to those sources, the higher the potential grade. The fewer of these that are used, and/or the more vague the references, the lower the potential grade.
Understanding of the material being presented.	This can be seen as the key criterion in qualitative assessment - **to what degree has the student transformed what they have found from the sources?** Little if any transformation of sources, description rather than analysis and interpretation, and listing rather than inter-relating or integrating would lead to a low grade ("P" or even "N"). A sound understanding and analysis, interpretation/inter-relating/integrating would lead potentially to an "H2B" grade, and a high level of abstract thinking, synthesis, originality, elegance, generalisation or application to other contexts would lead potentially to an "H1" grade.

O X-OZmAAA

University of Melbourne Institute of Education

APPENDIX D

B. Ed. (Sec.) Course

EDUCATION B/B1, 1993

ASSESSMENT FACE SHEET

"SCHOOLS AND THEIR FUNCTIONING" - FIRST ASSIGNMENT

Student's name:_____

Group:_____

CRITERIA See handout "Information for Students About Assessment and Grading Procedures".	**COMMENT**	**LEVEL ACHIEVED**				
		Nil	**Not satisf**	**Satisf**	**Very good**	**Exc.**
Relevance to question or task.						
Evidence of effort in preparation and presentation.						
Appropriate sources located and used.						
Understanding of the material being presented.						
Grade for this Assignment:		**N**		**P**	**H2B**	**H1**
				H3	**H2A**	
Staff member/date:						

DETERMINATION OF RANGES
FOR CONVERTING GRADE-NUMBER AVERAGES TO LETTER-GRADE

The process for combining the letter-grades for several pieces of work into a single final result is explained in detail in the section "Determination of Final Grade" on page 2 of the handout Information for Students About Assessment and Grading Procedures (Appendix A)., and in the handout Assessment and Grading - Guidelines for Staff (Appendix B).

This Appendix explains the calculations which determined the mark ranges shown in "Table C" and "Table 3", respectively, in those handouts.

Although the final grade classifications and their corresponding percentage-mark ranges (e.g., 80-100% = H1, 75-79% = H2A, and so on) were pre-specified by the University, it was still necessary for us to determine number-mark ranges to guide the process of converting the average of the grades on a student's individual pieces of work into a final grade for that student.

The first step was to set a number of clearly-recognisable "benchmark" levels. For example, a student who scored H2A on each of the six Education B assessment tasks should receive a final grade of H2A. Similarly, a student who scored H3 on each of the six pieces should receive H3 as a final grade.

These cases are simple and obvious, but what about the cases (more common) where students receive a mixture of grades over their several pieces of work? This is where the necessity for the temporary use of number-marks arises, and where boundaries or cut-off points between grade-levels become necessary

Take one of the "benchmark" cases - a student who receives an H2A grade for each of her six pieces of work. In number-mark terms, because H2A is equivalent to 5 marks on the 6 to 0 conversion scale, this converts to 6 pieces of work worth 5 marks each, which gives a total of 30 marks. Obviously, the average mark is 30 divided by 6, which gives 5, and 5 converts back to H2A as the final grade.

Using the same process, we can determine the minimum number-mark for a final H2A grade. It was decided that a student had to have at least half of her component grades at or above a particular level in order to receive a final grade at that particular level. For example, over six assessment tasks, 4 H2A's and 2 H2B's would earn a student a final grade of H2A because the majority of her work was at that level, but 3 H2A's and 3 H2B's would lead only to a final H2B grade.

Putting these examples in number-mark terms, the two calculations below give the following outcomes:

$$(4 \times H2A) + (2 \times H2B) = (4 \times 5) + (2 \times 4) = 20 + 8 = 28 = 4.66666 \text{ average}$$

$$(3 \times H2A) + (3 \times H2B) = (3 \times 5) + (3 \times 4) = 15 + 12 = 27 = 4.5 \text{ average}$$

Therefore, 4.66 could be taken as the lowest possible average to gain a student a final grade of H2A, while 4.5 would be the upper boundary for an H2B grade. This exercise was repeated for each of the grades H2B and H3 to determine upper and lower limits, but H1, P, and N needed slightly different treatment.

Clearly, it was not necessary to calculate an upper limit for H1 because it is not possible to score more than 36 marks in total, giving an average of 6. However a minimum limit needed to be calculated (5.6666).

A similar but reversed case existed with a P grade. It was necessary to calculate an upper limit (2.50) but the lower limit would clearly have to be 6 P's - an average of 2.00. Even if one grade of the six was below P, giving an average of less than 2.00, a final N grade would result.

Within the grade of N, we decided to have three levels, as described on page 10, above. Appropriate cut-off points were decided for each of these levels.

Having calculated upper and lower limits in this way on the basis of the six pieces of work involved in Education B, the exercise was repeated for the four pieces of work in the smaller subject Education B1.

Finally, to provide a broader picture to allow the number-mark ranges to be applied in a wide range of other situations, when necessary, the calculations were repeated so as to give ranges for any number of pieces of work between two and ten. This guided the rounding-off of the decimal fractions to the two places shown in the tables, and these mark-ranges can be used in any situation where there is a 0 to 6 scale and between two and ten separate pieces of work to be combined.

A similar process of determining upper and lower limits for each grade-level was used to determine the ranges to be used for the three graded component pieces of work within Education Studies C, as shown on the Assessment Face Sheets for that subject (for an example, see Appendix H - "Assessment Face Sheet for Assignment 1 - Classroom Data Analysis and Evaluation").

UNIVERSITY OF MELBOURNE INSTITUTE OF EDUCATION

APPENDIX F

MEMO

To: Education B/B1 Staff - JA, JB, BC, Marc D, Merryn D, ED, IG, KH, BH, TH, SL, DN, FO, BS, ES
From: Eileen Dethridge and Kevin Hall, Co-ordinators
SUBJECT: END-OF-YEAR RESULTS PROCEDURE - REMINDER
Date: November 3, 1993

Just a reminder about the process for determining final results in Ed. B/B1.

The full story is in the handout distributed earlier this year - "Assessment and Grading - Guidelines for Staff" (copy attached in case you have mislaid the first one), but if you haven't got a system already in place the table below might help in the collation of Assignment results. (You'll need one table for each student.)

Student's name:_____				Group:_____	
Assignment number	Assignment task	Letter grade given	Equivalent number mark	Weighting	Weighted mark
1 (2000 words)	Interviews/reflections			x 2	
2 (1000)	Observe/collect data about primary school (Ed. B) or local community (Ed. B1)			x 1	
3 (2000)	Teaching Area			x 2	
4 (1000)	Teachers' Work			x 1	
5 (2000)	Schools (fairness and parental choice, or educating all students) (Ed. B only)			x 2	
6 (1000)	Proposals for new school (Ed. B only)			x 1	
				Total weighted mark:	
				Final percentage mark: (Read from Table 2)	**%**

Remember that all components - 6 (or 4) Assignments, School Experience, attendance and participation - must be passed to pass the subject. That is, failure in one or more components will lead to failure in the subject even if the final percentage mark on the graded components is 55% or higher (see point 8 of "Assessment and Grading - Guidelines for Staff").

(EXTRACT FROM HANDOUT
"EDUCATION STUDIES C - GENERAL INFORMATION FOR STUDENTS")

The criteria below are general criteria that apply across all of the three graded assessment tasks in Education Studies C. They are supported by Specific Criteria for each of the three tasks that spell out in detail how these Basic Criteria apply to each of those particular tasks.

The Basic Criteria, and an explanation (in general terms) of what they mean, are shown below. They should be read in conjunction with the Specific Criteria for each task and the standard Assessment Face Sheet.

BASIC CRITERIA	GENERAL EXPLANATION
Completeness and relevance.	All parts of the requirements must be completed. Your response must be relevant to the question asked or task set. *A high score on this criterion should be easily achieved. by submitting what is asked for. A low score will result if some part or parts of the requirements are not completed, and/or if your response does not answer the question asked or meet the terms of the task set.*
Presentation and expression.	Presentation must be neat, clear, and legible. Spelling and expression must be literate. Reference to sources must use appropriate conventions. *A neat ,clear, legible and literate presentation will contribute to a high score on this criterion. (Artistic or other special presentation is welcomed but not expected). Untidy and/or unclear presentation, poor spelling, and/or poor expression will contribute to a low score, as will absence or lack of clarity of reference to sources.*
Location and use of sources.	Appropriate sources must be located. The sources must be used selectively. *Depending on the task, a wide or narrow range of sources may have to be used. The sources may be prescribed for you, or you may be expected to seek them out yourself.* *A high score on this criterion will result from locating the appropriate sources, and using them in a way that shows that you understand their meaning and significance for the argument or position you are presenting. A lower score will result if some or all of the expected sources are not used, and/or if your use of them does not demonstrate that you understand their meaning and significance.*

Understanding and transformation of the material being presented.	Raw data or other basic source material must be understood, and transformed in some way that develops it to a higher level. *A high score on this criterion will result from work which shows high levels of analysis and interpretation, inter-relating and integrating, abstract thinking and synthesis , originality, elegance, generalisation or application to other contexts, and in which all parts relate well to form a coherent whole.* *A low score will result if there is little if any transformation of sources, i.e., if there is description or listing rather than analysis and interpretation or inter-relating and integrating, and/or if the various pieces or stages of the total presentation do not link together well.*

A satisfactory score (i.e., 3 or above) must be achieved on each of the four Basic Criteria to be eligible for a "Pass" grade or above on that assessment task.

APPENDIX H

1994

UNIVERSITY OF MELBOURNE
INSTITUTE OF EDUCATION

B. Ed. (Secondary) Course
Education Studies C

Assessment face sheet for
ASSIGNMENT 1:
CLASSROOM DATA ANALYSIS AND EVALUATION

STUDENT'S NAME:_____

GROUP:_____ SEMINAR LEADER:_____

DATE SUBMITTED:_____

A grade may be amended up or down by no more than one grade level where it is felt that student's performance is not reflected appropriately by the number mark ("overall total"). Reasons for such amendments will be explained under "General comments".

Key to grade allocation:

H1	H2A	H2B	H3	P	N
16-15	14-13	12-11	10-9	8-7	6-0

STAFF MEMBER'S ASSESSMENT

Criteria	Level achieved					Comment
	Not shown	Not Satisfactory	Satisfactory	Very good	Excellent	
Completeness and relevance.	0	1	2	3	4	
Presentation and expression.	0	1	2	3	4	
Location and use of sources.	0	1	2	3	4	
Understanding and transformation of the material being presented.	0	1	2	3	4	

Column totals: + + + = (overall total)

General comment:

Staff member's signature:

Date:

The grade for this assessment task is:

UNIVERSITY OF MELBOURNE

INSTITUTE OF EDUCATION

Name: _____

Group: _____

B. Ed. (Secondary) Course

EDUCATION B, 1993

Assessment face sheet for:

ASSIGNMENT 5 - "SCHOOLS AND THEIR FUNCTIONING"

(Kevin's groups)

STAFF MEMBER'S ASSESSMENT							
Criteria *(See "Information for Students About Assessment and Grading Procedures".*	Comments	Level achieved					
		ns	Not satisf	Satisf	very good	Exc	
Relevance to question or task.							
Presentation and expression.							
Appropriate sources located and used.							
Understanding and transformation of the material being presented.							
	The grade for this Assignment is:	N		P	H2B	H1	
					H3	H2A	
Staff member's signature/date:	*General comment:*						

STUDENT'S ASSESSMENT						
Criteria *(See "Information for Students About Assessment and Grading Procedures".*	Comments	Level achieved				
		[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Relevance to question or task.						
Preparation and presentation.						
Appropriate sources located and used.						
Understanding and transformation of the material being presented.						
The grade I would give myself for this Assignment is:		N		P	H2B H3	H1 H2A
Date submitted:	*General comment:*					

APPENDIX J

ANALYSES OF STAFF AND STUDENT ASSESSMENTS

COMPARISONS ACROSS FOUR ASSIGNMENTS
(Education B/B1, 1993-94)

Table 1 Proportion of Students Participating in Co-Assessment

Table 2 Staff and Student Co-Assessments of Overall Assignment Grade

Table 3 Staff and Student Co-Assessments on Specific Criteria

Table 4 Co-Assessed Grade Distribution Compared With Staff-only Assessed Grade Distribution (Table 4A - 1993, Table 4B - 1994)

TABLE 1: ***PROPORTION OF STUDENTS PARTICIPATING IN CO-ASSESSMENT***

This table shows, for each of the four assignments and in total, the number and proportion of students participating in co-assessment.

Key figures are bolded.

	All students (n = 33 + 25 = 58)			**Females** (n = 27 + 17 = 44)				**Males** (n = 6 + 8 = 14)			
Assgnmnt no.	Total	Number who co-assd.	% who co-assd.	No. in class	% of class total	No. who co-assd.	% of total who co-assd.	No. in class	% of class total	No. who co-assd.	% of total who co-assd.
5/93	33	15	45.5	27	81.8	15	100.0	6	18.2	0	0
6/93	33	15	45.5	27	81.8	13	86.7	6	18.2	2	13.3
All cases '93	*66*	*30*	***45.5%***	*54*	***81.8%***	*28*	***93.3%***	*12*	*18.2%*	*2*	*6.7*
1/94	25	**2**	5.0	17	68	1	50.0	8	32	1	50
2/94	25	**5**	20.0	17	68	3	60.0	8	32	2	40.0
All cases '94	*50*	*7*	***14%***	*34*	***68%***	*4*	***57.1%***	*16*	*3 2%*	*3*	*42.9%*
All cases '93+'94	***116***	***37***	***31.9%***	*88*	*75.9%*	*32*	*86.5%*	*28*	*24.1%*	*5*	*13.5%*

TABLE 2: ***STAFF AND STUDENT CO-ASSESSMENTS OF OVERALL ASSIGNMENT GRADE***

All possible combinations of staff and student assessments are shown in the first column. (Staff assessment first, followed in brackets by student assessment.)

The shaded rows are the combinations where staff and student assessments agree.

The numbers in the cells show, for the overall grade for each of the four assignments, the actual number of cases of each possible combination.

Key figures are bolded.

Staff(Student) assessment combinations	Assignments				Totals		
	5/93	6/93	1/94	2/94	Females	Males	All
H1 (H1)	2	1			*3*		*3*
H1 (H2A)	2				*2*		*2*
H1 (H2B)	1	1		1	*3*		*3*
H1 (H3)							
H1 (P)							
H1 (N)							
H2A (H1)							
H2A (H2A)	1	1	1	1	*3*	*1*	***4***
H2A(H2B)	3	1		1	*4*	*1*	***5***
H2A(H3)	3	2		1	*5*	*1*	***6***
H2A(P)							
H2A (N)							
H2B (H1)							
H2B (H2A)		2			*2*		*2*
H2B (H2B)	1	4		1	*6*		***6***
H2B (H3)		1			*1*		*1*
H2B (P)	1	1			*1*	*1*	*2*
H2B (N)		1				*1*	*1*
H3 (H1)							
H3 (H2A)			1		*1*		*1*
H3 (H2B)							
H3 (H3)							
H3 (P)	1				*1*		*1*
H3 (N)							
P (H1)							
P (H2A)							
P (H2B)							
P (H3)							
P (P)							
P (N)							
N (H1)							
N (H2A)							
N (H2B)							
N (H3)							
N (P)							
N (N)							
Totals:	*15*	*15*	*2*	*5*	*32*	*5*	*37*

TABLE 3: ***STAFF AND STUDENT CO-ASSESSMENTS ON SPECIFIC CRITERIA (for each criterion on each assignment)***

All possible combinations of staff and student assessments are shown in the first column. (Staff assessment first, followed in brackets by student assessment.)

The shaded rows are the combinations where staff and student assessments agree.

The numbers in the cells show, for each assignment the actual number of cases of each possible combination for the assessments on specific criteria.

Key figures are bolded.

Staff(Student) assessment combinations	Criteria and Assignment Number																Totals		
	Relevance				Effort				Sources				Understanding						
	5/ 93	6/ 93	1/ 94	2/ 94	5/ 93	6/ 93	1/ 94	2/ 94	5/ 93	6/ 93	1/ 94	2/ 94	5/ 93	6/ 93	1/ 94	2/ 94	Fem	Male	All
Exc. (Exc.)	1	2		1	3	2		1	1	1			4	2			18		**18**
		4				***6***				***2***				***6***					
Exc. (VG)	7	7	2	3	5	2	1		6	2		1	4	4	2	1	42	5	**47**
Exc. (Sat.)	4	3		1	2	1						2	4	3			17	3	**20**
Exc. (Unsat.)														1				1	1
Exc. (Nil)																			
VG (Exc.)		1						1	2	1				2			7		7
VG (VG)	2				2	6	1	3	2	1	1	2	3	2		3	23	5	**28**
		2				***12***				***6***				***8***					
VG (Sat.)		2			1	3			3	1				1		1	8	4	12
VG (Unsat.)																			
VG (Nil)																			
Sat. (Exc.)																			
Sat. (VG)					1					6							7		7
Sat. (Sat.)	1				1	1			1		1						5		5
		1				***2***				***2***									
Sat. (Unsat.)																			
Sat. (Nil)																			
Unsat. (Exc.)																			
Unsat. (VG)																			
Unsat. (Sat.)										2							1	1	2
Unsat (Unsat.)										1								1	1
										1									
Unsat. (Nil)																			
Nil (Exc.)																			
Nil (VG)																			
Nil (Sat.)																			
Nil (Unsat.)																			
Nil (Nil)																			
Totals	15	15	2	5	15	15	2	5	15	15	2	5	15	15	2	5	128	20	148
Staff(Student) Agreement Totals		***7***				***20***				***11***				***14***			***46***	***6***	***52***

TABLE 4: *CO-ASSESSED GRADE DISTRIBUTION COMPARED WITH STAFF-ONLY ASSESSED GRADE DISTRIBUTION*

The two tables below show the number of grades at each level as finally recorded for co-assessed assignments and for staff-only assessed assignments.

1993 and 1994 data have been separated to highlight possible differences between the late-in-the-year 1993 assignments and the early-in-the-year 1994 assignments.

The numbers in the cells show, for the overall grade for each of the assignments, the actual numbers and percentages of cases at each grade level.

Key figures are bolded.

TABLE 4A: ***1993 assignments***

Grade level	Number at each grade level as finally recorded (n = 66)												
	Co-assessed (n = 30)							Staff-only assessed (n = 36)					
	5/ 93	6/ 93	F	M	*Total*	*% of co-assd.*	*% of class total*	5/ 93	6/ 93	F	M	*Total*	*% of staff assd.*	*% of class total*
H1	5	2	7		*7*	***23.3***	*10.6*	6	5	7	4	*11*	***30.6***	*16.7*
H2A	7	4	11		*11*	***36.7***	*16.7*	7	3	7	3	*10*	***27.8***	*15.2*
H2B	2	9	9	2	*11*	***36.7***	*16.7*	1	7	8		*8*	***22.2***	*12.1*
H3	1	0	1		*1*	*3.3*	*1.5*	4	3	4	3	*7*	*19.4*	*10.6*
P														
N														
Totals	15	15	*28*	*2*	*30*	*100%*	*45.5%*	18	18	*26*	*10*	*36*	*100%*	*54.6%*

TABLE 4B: ***1994 assignments***

Grade level	Number at each grade level as finally recorded (n = 50)												
	Co-assessed (n = 8)							Staff-only assessed (n = 42)					
	1/ 94	2/ 94	F	M	*Total*	*% of co-assd.*	*% of class total*	1/ 94	2/ 94	F	M	*Total*	*% of staff assd.*	*% of class total*
H1	1	1	2		*2*	***25.0***	*4.0*	3	5	7	1	*8*	***19.0***	*16.0*
H2A	2	3	3	2	*5*	***62.5***	*10.0*	6	1	5	2	*7*	***16.7***	*14.0*
H2B		1	1		*1*	***12.5***	*2.0*	6	10	12	4	*16*	***38.1***	*32.0*
H3								6	3	4	5	*9*	*21.4*	*18.0*
P								1	1		2	*2*	*4.8*	*4.0*
N														
Totals	3	5	6	2	*8*	*100%*	*16.0%*	22	20	28	14	*42*	*100%*	*84.0%*